

End of Term - 2020

Mark Phillips

University of North Texas Libraries

June 14, 2021

End of Term - History

- 2008 Bush -> Obama transition
- 2012 Obama -> Obama snapshot
- 2016 Obama -> Trump transition
- 2020 Trump -> Biden transition

Volunteer effort by contributing institutions

- Library of Congress
- Internet Archive
- UNT Libraries
- Stanford
- Environmental Data & Governance Initiative (EDGI)
- General Services Administration (GSA)
- Government Publishing Office (GPO)
- NARA

End of Term 2020

- URL Nomination Tool
 - Human contributed nominations
(<https://digital2.library.unt.edu/nomination/eth2020/>)
 - Bulk Lists
(https://digital2.library.unt.edu/nomination/eth2020_bulk/)
- Internet Archive – Wide Crawls
 - Pre Election - October 10, 2020 – January 19, 2021
 - Post Inauguration - March 18, 2021 – May 31, 2021
- UNT Libraries – Focused Crawls
 - Nominated URLs – December 14, 2020 – February 4, 2021



What was collected

EoT 2020 Crawl	Size	URLs	Hosts	.GOV	Start	End
Pre Election	97T	1.2B	996K	41%	2020-10-10	2021-01-19
Post Inauguration	192T	385M	702K	82%	2021-03-18	2021-05-31
UNT	3.3T	8.3M	99K	59%	2020-12-14	2021-02-04

Next Steps

- Access
 - Migrate eotarchive.cdlib.org site to new host.
 - Provide curated datasets to users to download
 - WARC, CDX, WAT?, WET?
 - Looking into different options, maybe based on Common Crawl datasets
 - End of Term - Github
 - <https://github.com/end-of-term>
- End of Term – Twitter
 - @eotarchive